

Gaze Tracking Enabled Automatic Robot Error Detection

Kaitlynn Pineda

CIS II Project Proposal

Mentors: Chien-Ming Huang and Maia Stiber

Introduction

In the field of human-robot interaction (HRI), researchers focus on how humans interact with robots and the different conclusions we can draw upon these interactions. HRI is a growing field with the potential for many applications in medicine or health care. Robots could assist with kitting in surgical settings, help patients in hospitals by providing bedside assistance, or even interact with COVID-19 patients in the ICU to help minimize the risk of exposure for healthcare workers. Nevertheless, robots are fallible and can make mistakes when executing their actions.

As the potential for robotic applications in the real world begins to grow, it is crucial to study how humans react to robot errors to gain a better understanding of humans' interactions with robots. Studying a human's physical and social responses to robot errors can inform researchers and roboticists of how not only robot errors may affect the trust humans place in a robot's actions, but also of ways to make future predictions of robot errors. If a robotic system does not have a built-in error detection system in place, recognizing that an error has occurred from a human's reaction is critical information; the feedback from such a detection system can allow a robot to minimize the severity of an error, allow a robot to correct its error, or even implement early stopping if the error is detected with enough notice.

There already has been a robot error detection system created, by my mentors on this project, that detects a robot error in real time by analyzing facial action units (AUs). Initial studies from my mentors show that people react very differently to physical mistakes executed by robots. Furthermore, depending on the task at hand and the users present, there may be a variety of human reactions. See Figure 1. The overall workflow of this current system consists of using two cameras aimed at a human's face, determining the current AUs present, detecting if an error has occurred based on these AUs, and logging the result with the corresponding timestamps. The ML algorithm that detects the error has been trained on a set of data of 19 participants reacting to robotic errors. The data collected was manually coded frame by frame in Microsoft PSI by two independent coders.



Figure 1: Sample reaction from Stiber pilot study. The image also shows the technical setup of the cameras used for AU detection.

Goals and Aims

The goal of this project is to build upon this system by introducing human eye gaze as another potential metric of robot error detection. The current AU detection system will still function correctly if a user wears glasses. Pupil Labs has released a mobile gaze tracker in the form of a pair of eyeglasses for a human to wear. We intend to have users wear these glasses while interacting with a robot as a robot error occurs to gain a better understanding of their gaze patterns. We hope to learn whether gaze can be an informative metric in robot error detection scenarios. Based on different types of errors, we want to investigate if human eye gaze is consistent; if there is a noticeable pattern that we could use down the line. Furthermore, we are interested in discovering if fixation points will appear that can inform us later in future scenarios explanations. Thus, the aims of this project consist of collecting data to understand human gaze reactions associated with physical robotic error, adding the gaze tracker data as an additional component to the Microsoft PSI system pipeline, and creating a ML algorithm with the data collected so that it will automatically detect the robot error as it occurs.

Materials and Methods

A user study was set up to gather human gaze reactions during physical robotic errors through a collaborative packing task.

User Study Design

Hypothesis: We can understand robot errors based on humans' reactions through gaze analysis.

We created an exploratory study design to analyze conceptual and physical robot errors during a human-robot collaborative packing task. We hoped to gather data on where people look during the task and where they look when the robot makes a mistake. While not a true factorial

design, the breakdown of our study design conditions is shown in Figure 2. We had four arrangements of two conditions.

Conceptual Task Error	Crush fragile object	Box won't close due to wrong object
Physical Task Error	Drop object	Doesn't pick up object
	Low Severity Error	High Severity Error

Figure 2: Exploratory Study Design; the task-specific errors used categorized by error type and severity.

Study Procedures

The participants were instructed to complete a packing task in collaboration with a Kinova robot. The participants were given a written set of instructions and the experimenter also verbally went over the instructions with them. The instructions clarified how the robot should be viewed as a packing expert, and what constituted a successful or unsuccessful packing task (i.e., no damage is done to the objects and the box can be closed without moving the objects within). They also explained the voice command pattern needed to correctly direct the robot to pack specific items. The participants were given a list of items that belong to each of the three boxes they were required to pack. The first box and its items remained the same, while the second and third boxes and items changed depending on the error and condition being tested. Thus, using voice commands, the participant instructed the robot which items to pack for each specific box. Each box had a set of three items the robot was required to pack. The planned error always occurred on the second item for the second and third boxes. The order of the conditions was randomly determined before the study, but each of the four error conditions was tested at least once. The experimental set up is shown in Figure 3.



Figure 3: Set up of the space where the user will complete the packing task in collaboration with the robot. The items for packing can be seen to the left of the robot, and the boxes used can be seen to the right of the robot. The list of items was given via the tablet on the table.

Participants

A total of 6 participants were convenience sampled for the pilot study. The participants' ages ranged from 22 to 38. All of them had engineering or computer science backgrounds and were somewhat familiar with robotics. The recruited participants were required to not need eyeglasses such that they could wear the gaze tracker glasses.

Measures

The data was recorded via the pupil labs invisible gaze tracker eyeglasses and a microphone. The invisible gaze tracker was automatically calibrated but recalibrated to each participants' gaze at the beginning of the study. While no paper questionnaire was used, the experimenters verbally asked the participants after the study if they witnessed an error, what the error was, and its corresponding severity.

The recorded data was downloaded from the Pupil Labs cloud software and manually annotated using the datavyu software, as shown in Figure 4.

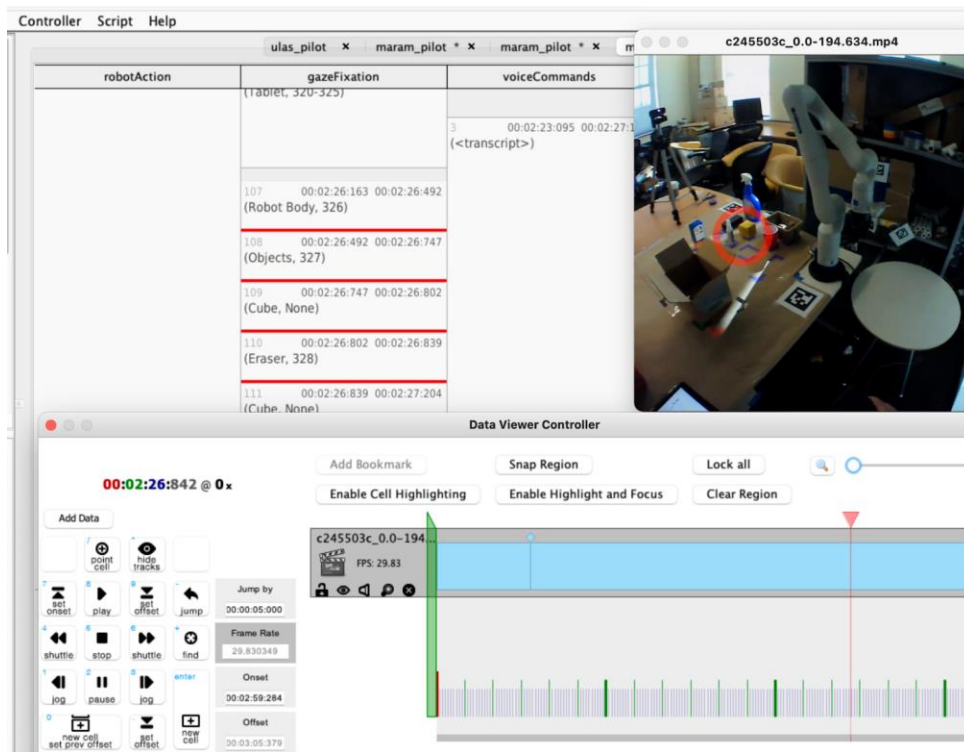


Figure 4: The datavyu coding software. The coders were able to label the corresponding object associated with a fixation, and the timestamps of gaze fixations with their corresponding fixation IDs provided by the pupil labs gaze tracker.

Gaze Fixations. The coder used the fixation circle provided by the pupil labs software to determine where the participant was looking. The coder then identified what object they were looking at during a timestamp range and labeled the timestamps accordingly to that object. They then matched up the corresponding time stamps to the fixation IDs with those associated timestamps in the data provided by pupil labs.

Results

While debriefing our participants, we learned that they did not always perceive some of the robot errors as an error. During the wrong object condition, they did not perceive it as a robot error that the wrong object was selected. It seemed to be more of an inherent error related to the study. Also, the error of the box not being able to close due to the selected object being too large once again was not perceived as a robot error, but rather an error by whoever requested those specific items to be packed for the box of that size.

In analyzing our manually coded video data, we confirmed that gaze appears to be goal oriented, as the literature indicates. Furthermore, gaze appeared to linger in an area where an error occurred (back at the object if the error involved missing the object, or at the gripper if the object was dropped). There also appeared to be occasional gaze shifts during the study. These mostly consisted of shifts back to the set of instructions or the tablet after an error occurred, or between the object involved in the error and the robot / gripper.

Since this analysis of results is qualitative, we decided our study could be improved if we implemented a more quantitative metric for gaze analysis to better categorize the gaze fixations. Thus, we began looking into how to measure gaze velocity during gaze shifts in real-time for our current workflow that only consisted of 2D gaze data with respect to the camera.

Challenges & Limitations

In order to measure gaze velocity, we needed to perform 3D localization of gaze within the world coordinate system. Thus, accurate detection of AR markers is essential for this process. However, we experienced several technical difficulties while learning how to measure gaze fixation velocity from the pupil labs gaze tracker. While first using AprilTags, we learned that they were not always detected during post-processing, and thus switched over to ArUco Markers. While these resulted in better detection than the AprilTags, we learned that there were still some issues with detecting the ArUco markers despite their size. We also encountered some inconsistencies in our transformation matrices for the markers/tags across images while using individual markers/tags. Thus, we transitioned to using an ArUco Grid Board detection rather than an individual ArUco marker detection. We are currently in the process of setting up the ArUco Grid Board in our study layout for automatic gaze velocity detection across fixations, and hope to conduct our future user studies using this technical approach.

Conclusion

While our qualitative analysis provided some insightful results, we realized it was essential to have a quantitative metric for gaze fixations. Due to the video coding being time consuming (at least 3 hours of coding per participant), we determined it was necessary to have a quantitative metric that could automatically provide additional information surrounding the fixation data in order to increase the scale of our user study.

Due to technical delays in calculating gaze velocity for the fixations, we have not yet integrated the gaze tracking component into the existing Microsoft PSI workflow. We hoped to run the gaze component independently in PSI first to determine how informative the gaze fixations are without the influence of AUs. Once our automatic gaze velocity measure is working and it is integrated into PSI, we can consider if implementing an ML algorithm to automatically detect the error would provide accurate results.

Our next steps, apart from finishing the physical setup for our new ArUco approach, involve finalizing our user study with errors that can be perceived more clearly and explicitly as errors by the participants (fixing the confusing objects by the participants and clarifying that the robot is a supposed packing expert). We also intend to include a full written questionnaire that will include questions regarding basic demographic information, and if the participants witnessed an error and its corresponding severity. We then hope to run a full-scale user study.

Acknowledgements

Thank you to Dr. Chien-Ming Huang for his advice with the project. Thank you to Maia Stiber for her help and guidance during our user studies, and for troubleshooting our technical issues.

Lessons Learned

Other than the knowledge presented in this report, I gained experience using AR tags to help perform the localization of the gaze tracker in world coordinates. I also learned that manual video coding is time consuming, and that thus quantitative metrics are preferred over qualitative metrics if possible, during human-robot interaction studies. In addition, with the help of my mentors, I learned how to design and run a pilot user study. While I have participated in and have helped facilitate human-robot user studies in the past, this was the first time I had designed one and conducted the initial data analysis portion of it.

Management

I met with my mentors, Chien-Ming Huang and Maia Stiber, weekly. I also communicated with them via Slack and email as needed.

	Original	Revised
Minimum:	<ul style="list-style-type: none"> • User Study Results and Analysis of Data • Integrate gaze tracker component into PSI for real-time error detection 	<ul style="list-style-type: none"> • User Study Results and Analysis of Data
Expected:	<ul style="list-style-type: none"> • User Study Results and Analysis of Data • ML Algorithm that automatically detects errors using gaze data 	<ul style="list-style-type: none"> • Integrate gaze tracker component into PSI for real-time error detection
Maximum:	<ul style="list-style-type: none"> • User Study Results and Analysis of Data • ML algorithm that classifies error based on severity OR second study to validate data collection 	<ul style="list-style-type: none"> • ML Algorithm that automatically detects errors using gaze data

Table 1: List of the deliverables for the project showing what the original ones were and how they were changed due to mentor recommendations and technical challenges encountered.

See Table 1 for a complete list of deliverables for the project. The results and discussion presented in this report meet the minimum deliverable and show the in-progress work for the expected deliverable. The revised expected and maximum deliverables were not completed for several reasons. Additional time was spent learning things that were not initially planned for due to unexpected technical challenges encountered after conducting our initial pilot study. There were also COVID-19 exposures that prevented us from conducting any additional in person pilot studies at the end of the semester. However, this project will be continued, and we hope to reach the expected and maximum deliverables in the coming weeks.

References

- Aronson, R. M., & Admoni, H. (2010). Gaze for Error Detection During Human-Robot Shared Manipulation. <https://www.semanticscholar.org/paper/Gaze-for-Error-Detection-During-Human-Robot-Shared-Aronson/390e9ac2e0a4162339a9a7e8bdb9c372bdcea659>
- Stiber, M., & Huang, C.-M. (2020). Not All Errors Are Created Equal: Exploring Human Responses to Robot Errors with Varying Severity. Companion Publication of the 2020 International Conference on Multimodal Interaction, 97–101. <https://doi.org/10.1145/3395035.3425245>
- Kontogiorgos, D., Tran, M., Gustafson, J., & Soleymani, M. (2021). A Systematic Cross-Corpus Analysis of Human Reactions to Robot Conversational Failures. In Proceedings of the 2021 International Conference on Multimodal Interaction (pp. 112–120). Association for Computing Machinery. <https://doi.org/10.1145/3462244.3479887>
- Honig, S., & Oron-Gilad, T. (2018). Understanding and Resolving Failures in Human-Robot Interaction: Literature Review and Model Development. *Frontiers in Psychology*, 9. <https://www.frontiersin.org/article/10.3389/fpsyg.2018.00861>
- Aliasghari, P., Ghafurian, M., Nehaniv, C. L., & Dautenhahn, K. (2021). Effect of Domestic Trainee Robots' Errors on Human Teachers' Trust. 2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN), 81–88. <https://doi.org/10.1109/RO-MAN50785.2021.9515510>